Weekly Report
11/16/2015 - 11/22/2015

Jing Xia

November 22, 2015

1 Summary

This week I mainly focus on the DataScanner project and idea initialization of VIS submission.

2 Projects

2.1 The DataScanner Project

I finished the experiment setting part of the user study section, which basically describes the goal, the hypothesis as well as the task. We divided the participants into two groups: the categorization tree group (CTG) and the small multiples group (SMG). Both groups were given the same system with data views, a matrix widget and relation viewers, except that the layout for CTG is the categorization tree while the layout for SMG is the small multiples. We first score users' categorization performance such that every correct pair of dimensions in the group is scored one. Then, an additional score is given if the relation of a correct pair is also correctly described. Finally, every participant received a total score and a relation description correctness rate. The score is the sum of both categorization scores and the relation description scores, which represents users' overall performance. The rate is defined as the relation description score to the categorization score, which indicates system's capability in presenting relation patterns.

In the first stage, we recruited only 14 participants (7 for CTG and 7 for SMG). Although from the data it seems CTG performs better than SMG, it has no statistical significance due to one or two bad performance. Thus I posted a recruitment on school forum in search of more participation. So far 24 participants responded, and 9 have completed the user study.

Besides the user study, I'm still investigating the mobile user case. Because the dataset is heavily skewed, I have some minor findings only. According to the data distribution, I am re-segmenting the data, expecting more in-depth findings afterwards. The new data is still under processing.

2.2 GeoScanner

The VIS submission project I want to do is on spatial-temporal data scanning. I intend to cover data cleaning, geo-spatial data calibration and data fusion. The geo-data type should include point-based, trajectory-based and region-based. However, after half week survey I realized that there are too many work to do to cover them all. Considering Zheng Yu's idea, I think visualization-aided data fusion might be a new perspective in visualization. But Zheng Yu's approach of knowledge fusion is more algorithm by algorithm and thus can hardly be general, which makes it unlikely to integrate the process with visualization. Only conventional data fusion (schema mapping) can be a proper topic for visualization. For schema mapping (mapping TAXI trajectory data with mobile cell data for example), the system should be able to 1. find the mapping key, 2. do mapping, 3. evaluate the mapping coverage (how much data is mapped) and 4. evaluate how mapping can complement one or both dataset. In this case, visualization is particularly helpful in 3. and 4., and can be of use with parameter setting or filtering in 2.. Another problem in geo-spatial data processing is about data matching (calibration). I think visualization can be of great help in extracting problems in data.
I will keep looking for more related work and have Zhengyang Shi start spark initialization first. The datasets include 1. TAXI trajectory data and mobile cell data for both calibration and fusion, 2. Purdue hurricane data (region-based) and Twitter data (point-based) for data fusion (not sure with this case). 3. other datasets mentioned in Zheng Yu's review paper.

3 Paper Reading and Miscellaneous

1 systems: worldlines,

- Waser, J., Fuchs, R., Ribiici, H., Schindler, B., Blschl, G., & Groller, M. E. (2010). World lines. Visualization and Computer Graphics, IEEE Transactions on, 16(6), 1458-1467. (to read)

2 spatial sampling: trifacta

3 data cleaning

4 data matching

- Lou, Yin, Chengyang Zhang, Yu Zheng, Xing Xie, Wei Wang, and Yan Huang. "Map-matching for low-sampling-rate GPS trajectories." In Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 352-361. ACM, 2009. (to read)
- Wang, Z., Lu, M., Yuan, X., Zhang, J., & Van De Wetering, H. (2013). Visual traffic jam analysis based on trajectory data. Visualization and Computer Graphics, IEEE Transactions on, 19(12), 2159-2168. (to read)

- Yuan, J., Zheng, Y., Zhang, C., Xie, X., & Sun, G. Z. (2010, May). An interactive-voting based map matching algorithm. In Proceedings of the 2010 Eleventh International Conference on Mobile Data Management (pp. 43-52). IEEE Computer Society. (to read)

5 data fusion

- Zheng, Yu. "Methodologies for Cross-Domain Data Fusion: An Overview." (2015). (to read)

6 spatialtemporal data mangement: location, trajectory, polygon

- location - Nanocubes **(done)**
- Z.L. Liu, B. Jiang, and J. Heer. imMens: Real-time visual querying of big data.Computer Graphics Forum (Proc. EuroVis), to appear, 2013. (to read)
- Gting, R. H., & Schneider, M. (2005). Moving objects databases. Elsevier. (to read)

7 ArcGIS API

8 GeoSpatial data uncertainty

- MacEachren, A. M., Robinson, A., Hopper, S., Gardner, S., Murray, R., Gahegan, M., & Hetzler, E. (2005). Visualizing geospatial information uncertainty: What we know and what we need to know. Cartography and Geographic Information Science, 32(3), 139-160. (to read)
- Kinkeldey, C., MacEachren, A. M., Riveiro, M.,& Schiewe, J. (2015). Evaluating the effect of visually represented geodata uncertainty on decision-making: systematic review, lessons learned, and recommendations. Cartography and Geographic Information Science, 1-21. (to read)
- Kinkeldey, C., MacEachren, A. M., & Schiewe, J. (2014). How to Assess Visual Communication of Uncertainty? A Systematic Review of Geospatial Uncertainty Visualisation User Studies. The Cartographic Journal, 51(4), 372-386. (to read)
- McKenzie, G., Hegarty, M., Barrett, T., & Goodchild, M. (2015). Assessing the effectiveness of different visualizations for judgments of positional uncertainty. International Journal of Geographical Information Science, 1-19. (to read)

9 Surveys, books

- Chen, Wei, Fangzhou Guo, and Fei-Yue Wang. "A Survey of Traffic Data Visualization." (2015). **(done)**
- Zheng, Yu. "Trajectory data mining: an overview." ACM Transactions on Intelligent Systems and Technology (TIST) 6, no. 3 (2015): 29. (to read)

- Andrienko, G., Andrienko, N., Bak, P., Keim, D., & Wrobel, S. (2013). Visual analytics of movement. Springer Science & Business Media. **(done)**

4 Miscellaneous

-

5 To Do List

1. DataScanner writing.
2. GeoScanner idea initialization.